IJTRET "Voice of Innovative Minds"

doi.org/10.54473/IJTRET.2021.5602

International Journal of Trendy Research in Engineering and Technology
Volume 5 Issue 6 December 2021
ISSN NO 2582-0958

Applications of Decision Trees- A Review

Harshitha M, S Bharath Prakash
Final year Under graduate students, Dept. Of Industrial Engineering And Management ,
RV College of Engineering , Bangalore,India
harshitham.im19@rvce.edu.in , sbharathp.im19@rvce.edu.in

Abstract

Decision trees are used at uncertain times to evaluate independent decisions. It contains decision points and decision alternatives. A decision tree is support apparatus with a tree-like construction that models plausible results, cost of resources, utilities, and potential outcomes using decision tree algorithms. This field of decision trees is aligned and more applicable for machine learning algorithms. This gives the outcomes of the decision path by the decision parameters considered and the problem is taken into account. Here, the multiple areas are discussed discuss marking the applications of the decision trees in the field of- production processes, mining, pharmaceutical processes, manufacturing processes, Biotechnology, and medical diagnosis.

Keywords: Decision trees, production processes, data mining approach, biomedical, algorithms.

1. Introduction

Nowadays besides theimprovement of the general interaction execution, the support of safe workingconditions is the critical component in the improvement of interaction control frameworks In the process industries, the increasingly complex technologies affect considerable challenges in their design, analysis, manufacturing, and management for successful operation. For many applications, it is necessary that maintain the process variables within strict limits. Due to theintricacy of creation frameworks the issue counteraction,analysisthe control of strange occasions turned out to be increasingly more muddled for the cycleadministrators. A huge number of interaction factors are noticed and put away everysecondin a huge interaction plant. To support the operators in improving the quality of products, reducing the energy and materials waste, and increasing the flexibility of production it is necessary to increase their insight into the behavior of the process. Due to this demand, the number of claims against developing diagnostic systems is continuously increasing. Trees have been a very widely used approach for classification owing to their advantages which include smooth handling of various types of features such as numerical and categorical, the capability of capturing nonlinear and non-additive relationships between features while they remain easily interpretable.

2.Applications of Decision trees

2.1 Decision trees in process control systems- a division under production processes.

Process performance plays a crucial function of any product to improvise its quality, reduce the usage of energy and excess of materials along with proper maintenance of process operators. Minimization of false operations and failures are essential to reduce production breakdowns. Hence, this new emphasis on Decision trees aids the process control systems to innovate and alienate in the case of the dynamic process as a focus by using a heterocatalytic reactor[1].

On-line diagnostic and fault detection gives the source of information between calculated and available data during the construction of an integrated process model which can be understood by the following scheme.

International Journal of Trendy Research in Engineering and Technology
Volume 5 Issue 6 December 2021
ISSN NO 2582-0958

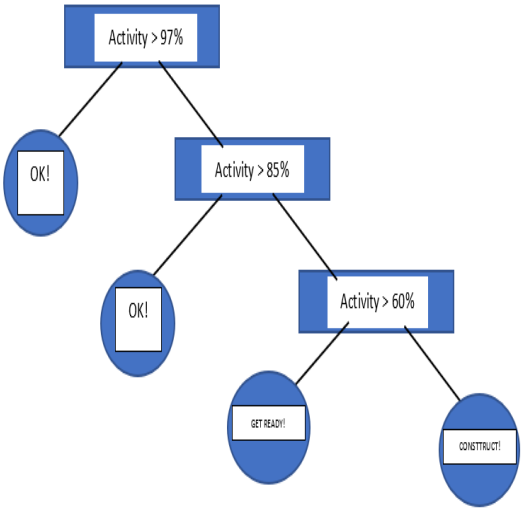

Fig.1. The worked-out decision forest

Thus, it can be carried out by the binary decision trees with the help of classifiers and 2 types of nodes: (1) internal node with 2 children (2) external node without children. As mentioned, the heterocatalytic reactor aids to focus on two important phenomena- reactor runaway and catalyst aging. This work establishes the connection between decision trees and process control systems of dynamic processes with the stability-instability caused by the usage of heterocatalytic reactor. The below decision tree summarizes the outcomes.

2.2 Decision trees-based knowledge mining approach in complex production systems

In this case of application, the decision trees are constructed using one of the machine learning algorithms- Reinforcement Learning Algorithm (RLA) to obtain predictive modeling in Data mining. This algorithm inputs 2 main sub algorithms: LMT and Hoeffding Tree Algorithm with comparing trees and results having an intermediate stage to extract rules to complete the process. The combination of production, maintenance, and recycling policies was derived by the main Algorithm- RLA[2].

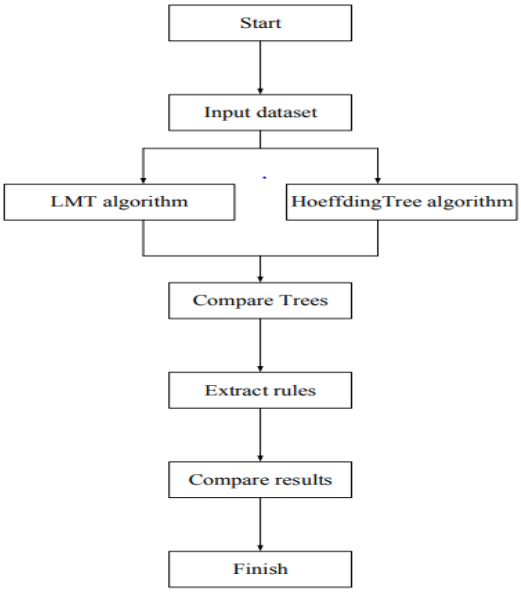

Fig. 2. The flowchart of the proposed approach

After data set loading, the processing output of both algorithms is summarized below.

Table. 1. Comparative results for the algorithms of the analysis

FEATURES	LMT ALGORITHM	HOEFFDING TREE ALGORITHM
MEAN ABSOLUTE ERROR	0.0236	0.0617
RELATIVE ABSOLUTE ERROR	8.31	21.72
TIME CONSUMED	1.94 sec	0.05 sec
ACCURACY	96.23	89.43
SPEED	SLOW	FAST
PERFORMANCE	BETTER	GOOD

IJTRET "Voice of Innovative Minds"
International Journal of Trendy Research in Engineering and Technology
Volume 5 Issue 6 December 2021
ISSN NO 2582-0958

2.3Decision tree in pharmaceutical production

The data sheets contain multiple factors if the microorganisms were objectionable or not which is used to analyze and determine throughout the entire process of pharmaceutical production. With the aid of decision trees, the evaluation was conducted to perform on 3 main variant factors viz., product-related, hygiene-related, and recipient-related with isolation of the microorganisms as the starting process. This decision-making tool was compatible with a systematic procedure involving making quick and easy decisions along with verification and evaluation as depicted below[3].

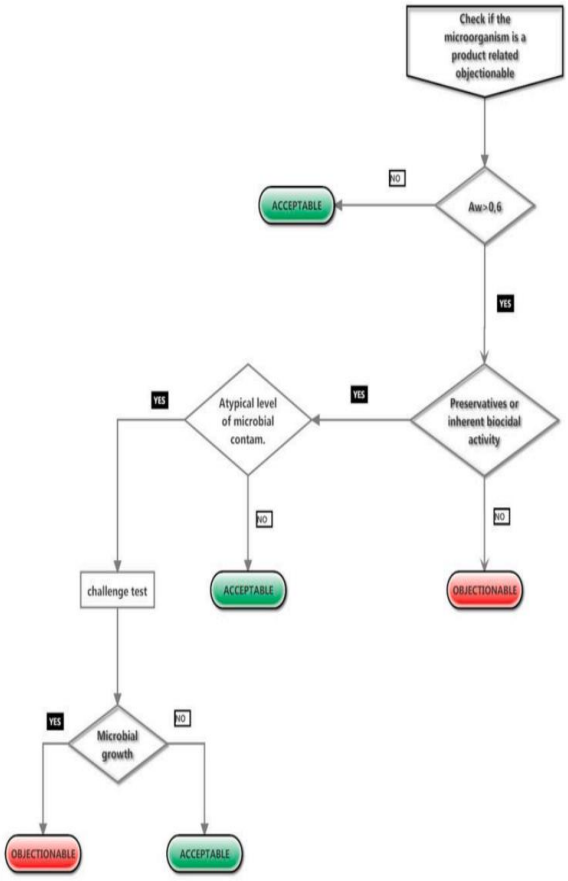

Fig. 3. Decision Tree Flowchart- evaluation: Is the Microorganism product-related objectionable?

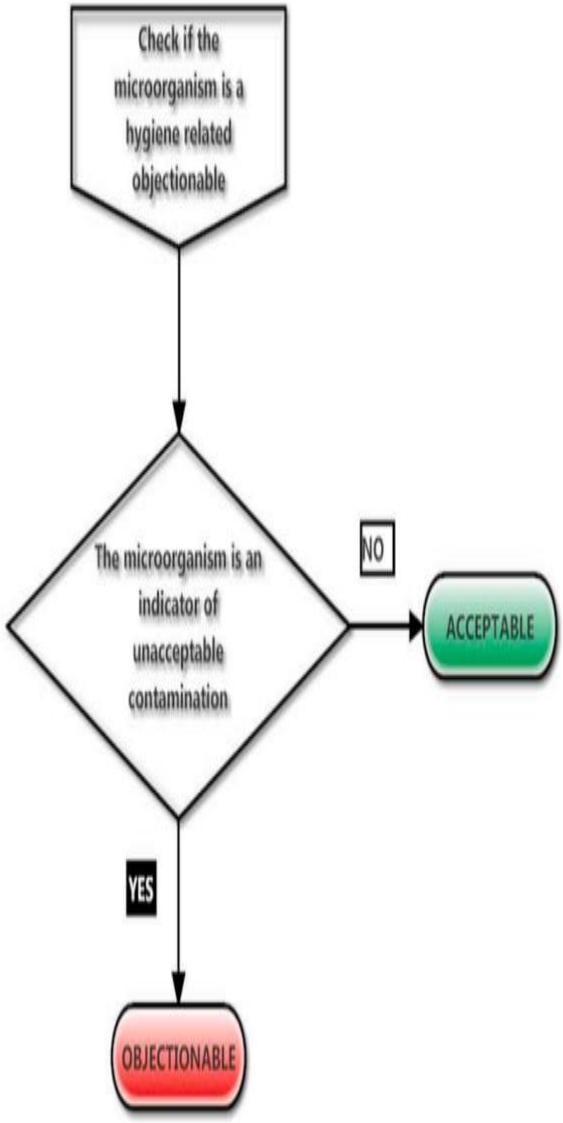

Fig. 4. Decision Tree Flowchart- evaluation: Is the Microorganism hygiene-related objectionable?

International Journal of Trendy Research in Engineering and Technology
Volume 5 Issue 6 December 2021
ISSN NO 2582-0958

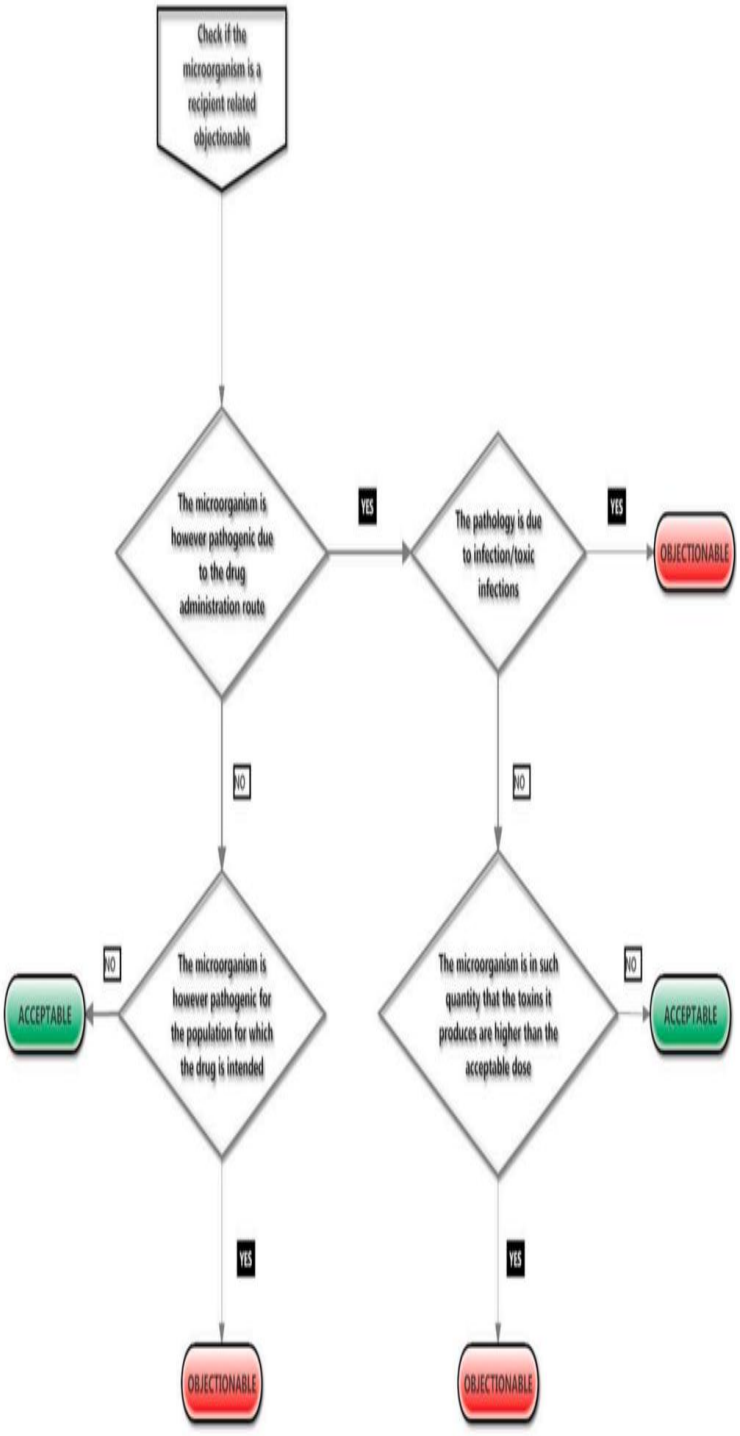

Fig.5. Decision Tree Flowchart- evaluation: Is the Microorganism recipient-related objectionable?

The data sheets had several microorganisms belonging to different taxa and phylum amongst which the different factors are filtered which is causing objectionable in pharmaceutical production. In this, it was analyzed by suitable software, compliant with the Code of Federal Regulations (US Food and Drug Administration, 2003), allowing the storage of the collected data and their treatment keeping the decision tree as a center of the focus.

2.4Decision tree approach for Identifying Defective Products in the Manufacturing Process

There have been several approaches proposed to improve the efficiency of the manufacturing process using data mining techniques. By using the data mining technique, the authors create an intelligent tool for extracting useful information automatically which enables the engineers and the managers to understand the complex manufacturing data easily[4].

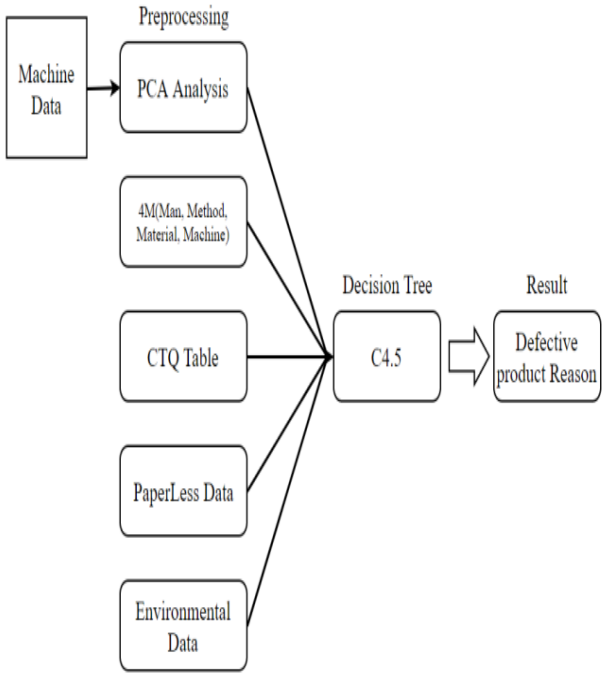

Fig.6. A conceptual design of the proposed method.

2.4.1 Dataset:

4M data is collected through the existing MES, POP, and ERP and the CTQ table used the values of the allowable range. This range is defined by the manufacturer based on the attributes that affect the quality of each product. PaperLess refers to the data collected by constructing a computerized system for process management. PaperLess is measured twice a day (day and night), and the measured value is displayed along with the occupant's number when entered. Environmental data is collected in the unit of temperature/humidity in the process, temperature/humidity sensor inside the warehouse, and temperature/humidity provided by the meteorological office.

2.4.2 Data Analysis:

The algorithm C4.5 of Decision Tree uses the method of dividing by entropy index and information gain.

$$Entropy(S) = \sum_{i=1}^{m} p_i log_2(p_i)$$

$$p_i = \frac{freq(C_i, S)}{|S|}$$

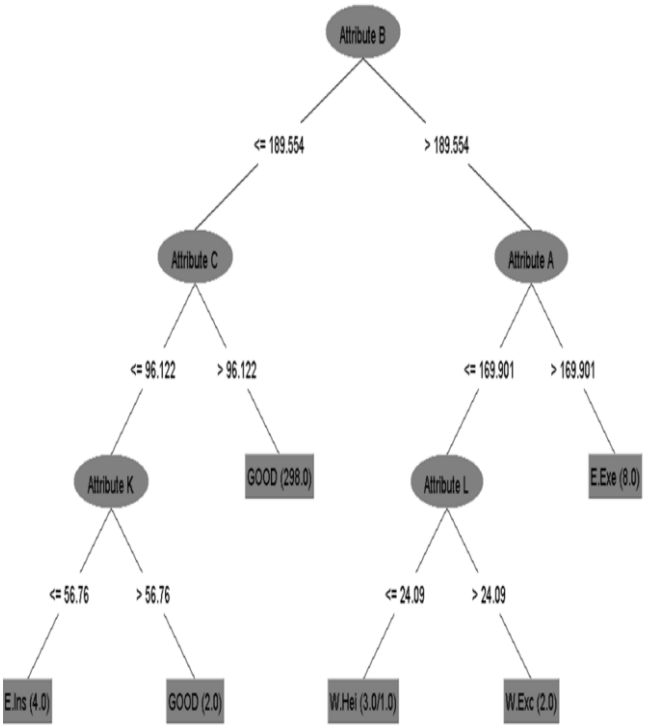

Fig. 7. Decision tree

2.4.3 Performance evaluation:

Summarizes classification accuracy according to the amount of data. When the number of data is 50, and the ratio of good and bad is about 8:2, classification accuracy is about 95%. It can be said that about 2.5 out of 50 cases cannot be categorized. When the number of data is 100, and the ratio of good and bad is about 8:2, classification accuracy is about 95%. Finally, if the number of data is 300 and the number of good products is 263, the classification accuracy is 96%, which is the highest value. For the training data of Process, a decision tree model was created based on 890 good data and 110 bad data out of a total of 1000 data.

Table 2. Performance Evaluation of the decision tree

Number of data	Number of Products	Number of bad Products	Accuracy
50	39	11	95.24%
100	81	19	95.32%
300	263	37	96.01%

Based on the data collected in the process, we have proposed to investigate the causes of defective products using the C4.5 algorithm of decision trees. The existing methods to increase the productivity in the manufacturing process did not show a great effect due to the lack of work I experience and the single process-oriented analysis. In addition, most of the existing researches has been conducted to compare the results of the analysis according to the size of the data.

2.5 Process Optimization by Molecule Swarm Optimization and Gradient Boosting Decision Tree Algorithm

The creationof reasonable clean energy can be accomplished by co-pyrolysis of farming deposits andwastewater ooze. The non-added substance warm conductof co-pyrolysis of drug muck and Ginkgo Biloba leaf buildups was explored.Thesynergistic impact of co-pyrolysis wasnot clear at raised temperatures. Further, dynamicthecs of co-pyrolysis was consideredby fittingtheCoats-Redfern incorporation technique to thethermogravimetric (TG) bend.The difference in hotness and mass exchange in the reactor caused the difference in powerfulboundaries. Additionally,themixture molecule swarm enhancement and slope helpingchoice tree (PSO–GBDT) calculation was intended to help the energy creation atfull-scale pyrolysisplantsby checking the TG curve.[5]

PSO–GBDTmodel well predicts mass misfortune pace of the blend at various warming ratesaffirming that co-pyrolysis of PS and GBLR can resultin high energy creation byexpanding PS pyrolysis. Planning PSO–GBDT model assistance to decreased waste creationby ingenious treatment of wasteinto energy. Thermogravimetric examinationof GBLR andperson. The pyrolysis cycle of PS could be recognized into three-personstages: Pre-warming stage (mass lost 4.73% under 138 °C), unstable devolatilization stage(mass lost 46.59% at 138–572 °C), and carbonization stage (mass lost 4.22% at 572–800 °C).

Initially, the preheating stage related to the arrival of an exceptionally low amount ofvolatiles and drying out of natural substances by loss of both free water and chemicallybondedwater.Since the natural substances utilized in the analysis had been dried andthe water content was low, the varieties in the TG bends were little. The heat wasabsorbed during the vanishing of water. The vanishing pace of water came to amaximum at roughly66°C,Furthermore, a pinnacle showed up in the DTG results, which wascaused by the solid precipitation ofwater.

At roughly 138°C,the water had essentially vanished, and a modest quantity ofdepolymerization and"glass change"happened inside the unrefined substance at this stage.The unpredictable devolatilization stage was the fundamental phase of pyrolysis of unrefined substances.At this stage, the substance obligations of lipids, proteins, and sugars in the PS were broken, anenormous number of little atom gas-stage parts and fluid stage partswith a moderately huge atomic weight were shaped, and a lot of tar wasaccelerated. The nature of the unrefined components was fundamentally decreased during this stage,also the TG bend was extraordinarily diminished.The DTG bend originally had a conspicuous weight reductiontop at this stage, and a shoulder top showed up behindit.Thejustification for this outcome isthat the unpredictable parts in PS are generally muddled and the strength of thesubstance bonds among the parts is conflicting. The weight reduction and pyrolysispace of the

example arrived at the most extreme at thisstage.Themost extreme weight reduction toptemperature was340°C,what's more, the most extreme weight reduction rate was 4.9%/min. Thecarbonization stage included the sluggish decay of the buildup.

Sincethe interaction by and large went on for quite a while, the precipitation of biomass volatileswas exceptionally lethargic, so the TG bends and DTG bends would in general be delicate.If there should be an occurrence of GBLR,there were likewise four phases associated with the debasement of GBLR: pre-warming stage(mass lost 5.53%below126°C),unstable devolatilization stage (mass lost 58.68% at126–552°C)what's more carbonization stage (mass lost 5.43% at 552–800 °C). The main stagewas like PS, which was brought about by parchedness and water vanishing. The unpredictabledevolatilization stage was the fundamental pyrolysis phase of the GBLR.

At this time, the GBLR delivered an enormous weight reduction top because of precipitation of the lossof a lot of volatiles due to cellulose and hemicellulose deterioration.Hemicellulose is the most unsound and most effectively pyrolyzed among the threeparts. Hemicellulose, cellulose, and lignin wereeffectivelydisintegrated, yet no conspicuous temperature limit was noted among the three pyrolyticprocesses .

Numerousfurious pyrolysis responses happened in this stage, then, at that point, a lot of unpredictablematters were created, which were made out of little sub-atomic weight gases, for example,hydrocarbons,CO2, CH4,and so forth . The third stage was the burninginteraction of a limited quantity of fixed carbon; chiefly, the high-edge of boiling over natural matterin GBLR disintegrated, a pinnacle showed up in the DTG bend, and the most extreme weight reductiontop temperature and greatest weight reduction rate were 346 °C and 7.81%/min,separately. Notwithstanding, the response at thisstage was extremely sluggish, and the interactionenduredfor a longtime.

Subsequently, the varietyin the TG bend wasextremely delicateSynergistic impacts between theGBLR and PSTo completely comprehend the synergistic impacts happening during theco-pyrolysis interaction of GBLR and PS, the hypothetical TG worth of the combination wasdetermined dependent on the proportion of the two examples and their TG esteems under something very similartrial conditions. The hypothetical incentive for the pyrolysis of the combination wasacquired by ascertaining the weighted normal amount of the trial upsides of theindividual examples, which can be acquired by the accompanying techniques.

$$W_{\text{Calculated}} = X_{\text{G}}W_{\text{G}} + X_{\text{P}}W_{\text{P}}$$

where X_{G} and X_{P} are the blending ratio of the GBLR and the PS, and under the same

conditions, W_{G} and W_{P} represent the weight loss values during the pyrolysis of GBLR and PS respectively. The synergistic effects can be described by the discrepancy (ΔW) between experimental and calculated weight loss values as follows:

$$W = W_{\text{Experimental}} - W_{\text{Calculate}}$$

where $W_{\text{Experimental}}$ and $W_{\text{Calculated}}$ represent the experimental and theoretical weight loss values. It can be concluded that if ΔW is greater than zero, it means that the released volatiles are more than expected.As announced in past research, the synergistic system during the co-pyrolysis processof the mixesis fundamentally ascribed to the arrival of hydrogen and hydroxy extremist andthe reactant impacts of antacid and soluble earth metals in materials.There is a synergisticcollaboration between the mixes.The above outcomes demonstrate that abundance volatiles in theblend negatively affect co-pyrolysis.

Kinetics analysis

The Coats-Redfern method was used to calculate the pyrolysis kinetic parameters E and A and the correlation coefficient R^2 at different heating rates β. The kinetic parameters were calculated at

heating rates of 10, 20, 30and 40 K/min. The straightrelationship coefficientsof the situation fitted by the two-layered compound response and the Avrami-Erofeev (n=1)model were moderately high (all above 0.98). Theinitiation energy E and recurrence factor A were distinctive at various warming rates. Theinitiation energy of the blend expanded with expanding warming rate, and theinitiation energy expanded from 33.94 kJmol to 35.25 kJ/ mol.Wipe outthe impacts of trial investigation mistakes. This was because the higher thewarming rate, the more prominent the hotness move opposition inside the reactant particles, andthe higher the remotely required energy level for the response. Interim, the incrementin initiation energy was joined by an expansion in the recurrence factor, whichwas essentially brought about by test conditions like test gear and warming. The difference in hotness and mass exchange in the reactor caused the difference in powerfulboundaries.

To this end, the E and InA of the combination at various warming rateswerestraightly fitted.In this way, the effect of the frequency factor on the change of the activation energy can be partially compensated, and the kinetic parameters after the compensation process are much less affected by the experimental conditions. The linear correlation coefficient R^2 of the fitted equation was 0.9798, and the dynamics compensation effect expression of the mixture was ln*A* = 1.1186 8*E*- 33.95134. The *E* and *A* were substituted into Eq. 1 to establish the kinetic equation for the mixture pyrolytic process.

In the GBDT model, GBDT is an integrated

$$\frac{d\alpha}{dT}=\left(\frac{e^{(1.1186E-33.95134)}}{\beta}\right)\exp\left(\frac{E}{RT}\right).\frac{(1-\alpha)[-\ln(1-\alpha)]^2}{3}$$

algorithm based on decision trees. To encounter, the principle boundaries influencing the improvement capacity ofthe GBDT model to incorporate the learning rate, the number of choice trees, and themost extreme tree profundity requirement for every choice tree. The three boundaries connect. It is hard to change the other two boundaries physically to settlethe boundaries to get the ideal boundaries. The three boundaries connect, so it is hard to acquire the most boundaries by settling one boundaryalso changing the other two boundaries. The conventional Gird searchCV technique requires a parcel ofit.

Itshould be visible that later the original, in the underlying 30 populaces, the greatestfitting degree was 0.9853, and later 100 ages, it at long last rose to0.9887.Forexamination, the 5-overlay cross-approval normal of the GBDT model utilizing the defaultboundaries was0.9797. The fitting level of the PSO-GBDTmodel on the test information was 0.9987, and the MAPE was0.2272%.Because the fitting degree of these two models is relatively high, the difference between the two models cannot be seen intuitively. However, the GBDT's MAPE is 0.5473%, and the PSO-GBDT reduces this error by more than 50%. Therefore,compared with the GBDT model with default parameters, the performance of the PSO-GBDT model is greatly improved. From the fitting graph, it can be seen more intuitively that the test points of PSO-GBDT fit the 45-degree fitting line more. In the process of gradually decreasing the true value of GBDT, the error gradually becomes larger, especially in the interval of [-0.25, -0.15], and many points have obvious deviations.

Hence thermal and kinetics of co-pyrolysis of GBLR and PS were studied. Major deviation in TG curve of pyrolysis could be observed at medium-temperature (200– 500 °C) than that of high-temperature. The synergistic effect of co-pyrolysis is mainly attributed to the release of hydrogen and hydroxy radical and the catalytic effects of alkali and alkaline earth metals in materials. Variouskinetic parameters and TG data in co-pyrolysis can be observed usingprediction data obtained bythe PSO–GBDT model. Conclusively,co-pyrolysis of the twomaterials will resultin a reduction in thetoxicity of waste by converting themto sustainable energy products

2.6 Decision Tree Models for Medical Diagnosis

Theheartdisease dataset, diabetes dataset, and liver disease dataset fromthe UCI machininglearning repository are used forclassificationtasks.66computational overheads.into the% of the dataset is utilized for preparing and staying 34 % is utilized for testing. The heartinfection dataset contains 270 perceptions and 2 classes: the presence and nonattendance ofcoronary illness.There are 150patient records without suffering.The results of classifiers are shownbelow[6].

Table 3. Prediction results of heart disease dataset

	ASTree	J48	NBTree	Random Forest	Random tree
Exactly predicted	40	44	45	46	44
Not exactly predicted	13	9	10	7	10
Accuracy	75.52%	80.2%	85.1%	87.8%	82.79%

The diabetes dataset contains 768 occurrences and 2 classes:the presence and nonappearance of diabetes. There are 500 patient records without enduringdiabetes and 268 records for a patients with diabetes.

Hepatitis illness dataset contains 155 occasions and 2 classes: expressing the existence visualizationindeed (or)no. There are 123 patient records for life anticipation yes and 32 records fora patient with no.

Similarly, the results were summarized for the other two cases.

3.Discussions and Conclusions:

In the present paper, we perform choice trees calculationsto build trees for addressing approaches found as best by a

Support Learning-based calculation while treating a creation controladvancement issue. The choice tree is built to outlineChoice tree calculation,theelements ofnumerous factors with discrete qualities. The leaves of the tree relate to the arrangement ofwork esteemswhat's more the non-terminal hubs to its autonomous factors. Giventhat current investigations utilize dynamic programming for treating such issues, thecommitment of the review is the use of notable order calculationsextricating choice guidelines concerning a creation control framework, from a dataset madea calculation. These standards can be altered to different setups of some othercreation framework. The subsequent trees outline that the various calculations can makevarious trees with significant contrasts in quality. This examination can bestretched out by applying more proficient calculations and making a broad relativecomputationalinvestigation for deciding the best-performing calculation for the givenissue.

References:

[1]Applying decision trees to investigate the operating regimes of a production process T. Varga, J. Abonyi, F. Szeifert University of Pannonia, Faculty of Engineering, Department of Process Engineering, H-8200 Veszprém, Egyetemút 10- 2007

[2] Athens, Greece. A Decision Trees-based knowledge mining approach for controlling a complex production system- 2021

[3] Objectionable microorganisms in pharmaceutical production: Validation of a decision tree Susi Burgalassi a, Stefano Ceccanti b, Sandra Vecchiani b, Giulia Leonangeli b, Ileana Federigic,*, Annalaura Carducci c, Marco Verani c

[4] Sungsu Choi R&D Center YURA Co., Ltd., Seongnam, Gyeunggi, South Korea: A Decision Tree Approach for Identifying Defective Products in the Manufacturing Process- 2017.

[5] Efficient Pyrolysis of Ginkgo Biloba Leaf Residue and Pharmaceutical Sludge (mixture) with High production of Clean Energy: Process Optimization by Particle Swarm Optimization and Gradient Boosting Decision Tree Algorithm:Zhenwei Yu, Khurram Yousaf, Muhammad Ahmad, Maryam Yousaf Methods, Qi Gao, Kunjie Chen, 2020.

[6] Decision Tree Models for Medical Diagnosis: Aung NwayOo, Thin Naing- 2019.